Metadata management system в Авито

Крючков Фрол







Фрол Крючков

В Авито с 2015 года

— Занимался backend-разработкой

— Team lead@core-services

— Team lead@trust&safety

Об Авито

Всего **сотрудников**:	**2200 + чел.**
Команда **разработки**:	**650 + чел.**
Средний **возраст** сотрудников:	**28 лет**
Количество **серверов:**	**1000+**
RPS к бэкенду:	**96 000**
MAU:	**50 миллионов чел.**
Релизов изменений:	**200-250 в день**
Команд разработки:	**58**
Модерация:	**4 миллиона версий объявлений в день**
Продано товаров в 2019 году**:**	**1,32 триллиона рублей, или, 1,45% ВВП России**



План

Что такое **metadata**

Почему важно этим заниматься

Ограничения старой архитектуры

Архитектура **metadata management system**

Deep dive сервиса **layout**

Плюсы/минусы



План



Что такое
metadata



metadata — это данные о данных



Descriptive metadata

TITLE ...

CATEGORY CARS

PRICE $100

DESCRIPTION

...



Structural metadata

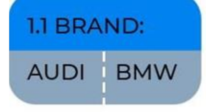
SUBMITTING A NEW ADVERTISEMENT

1.1 BRAND:	
AUDI	BMW

1.2 BRAND:	AUDI
MODEL	A1
	A2
	A3
	A4

1.3 BRAND:	AUDI
MODEL	A3
YEAR	2020
	2019
	2018

SEARCHING



1.2 BRAND:	AUDI
MODEL	A3

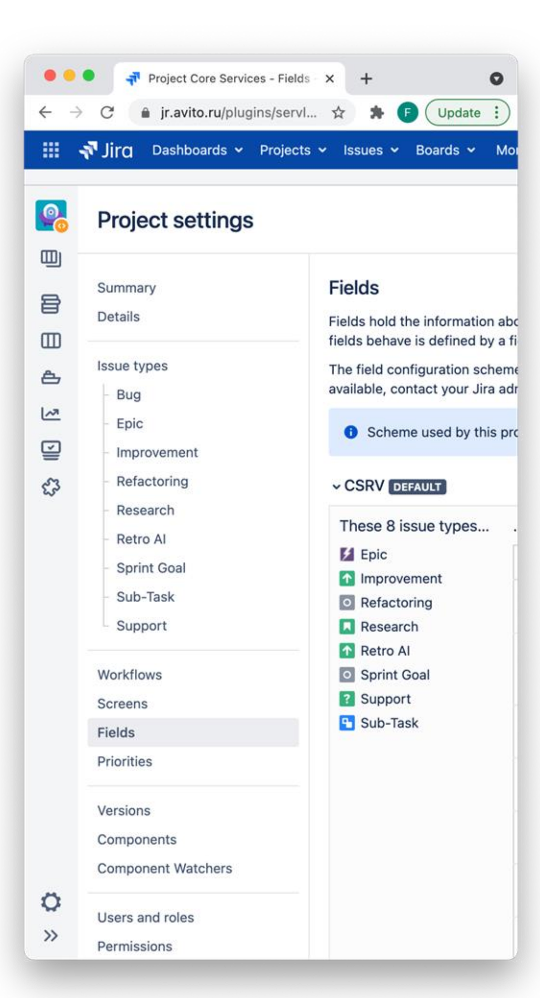
1.3 BRAND:	AUDI
MODEL	A3
GENERATION	2018-2020
	2016-2017



Какие решения уже существуют?



Какие решения уже существуют?





Какие решения уже существуют?

Manage Categories ‹ Site Title

wordpress.com/settings/taxonomies/category/frolsblog.wordpress.com

My Site Reader Write

Site Title
frolsblog.wordpress.com
Coming Soon

Free domain with an annual plan Upgrade

- My Home
- Stats
- Upgrades Free
- Posts
  - All Posts
  - Add New
  - Categories
  - Tags
- Media
- Pages
- Comments
- Feedback
- Jetpack
- Appearance
- Plugins
- Users

← Back Categories

Search... Add New Category

Uncategorized DEFAULT 3

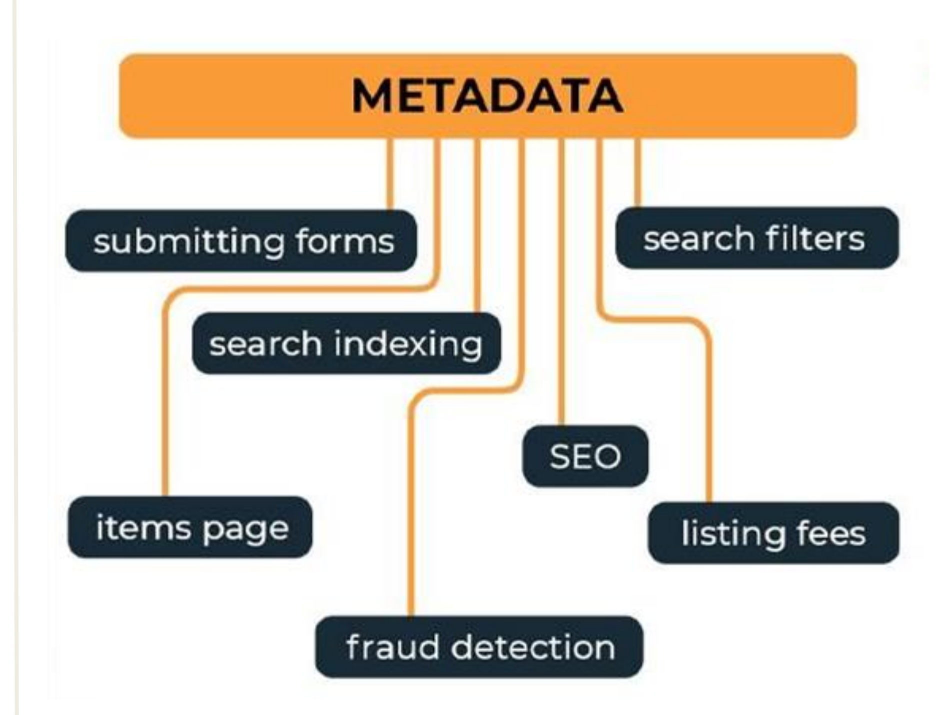


План

Почему важно
этим
заниматься



Роль metadata для сайта объявлений





Процесс изменения метаданных

- UI изменения форм
- Изменения валидаторов
- Антифрод модели
- Рекомендательные системы
- Ценовые политики
- Индексация
- SEO



План

Ограничения
старой
архитектуры



Проблема

- UI изменения форм
- Изменения валидаторов
- Антифрод модели
- Рекомендательные системы
- Ценовые политики
- Индексация
- SEO

От одной недели до нескольких месяцев



Functional requirements

- Уменьшить время внесения изменений
- Разное представление на разных платформах
- Поддержка a/b тестирования
- Поддержка версионирования



Старая архитектура

ID	TITLE	CATEGORY ID	PARAMETER 1	...	PARAMETER N	PRICE	DESCRIPTION
1	...	...	...	...	...	...	...
2	...	...	...	...	...	...	...
3	...	...	...	...	...	...	...



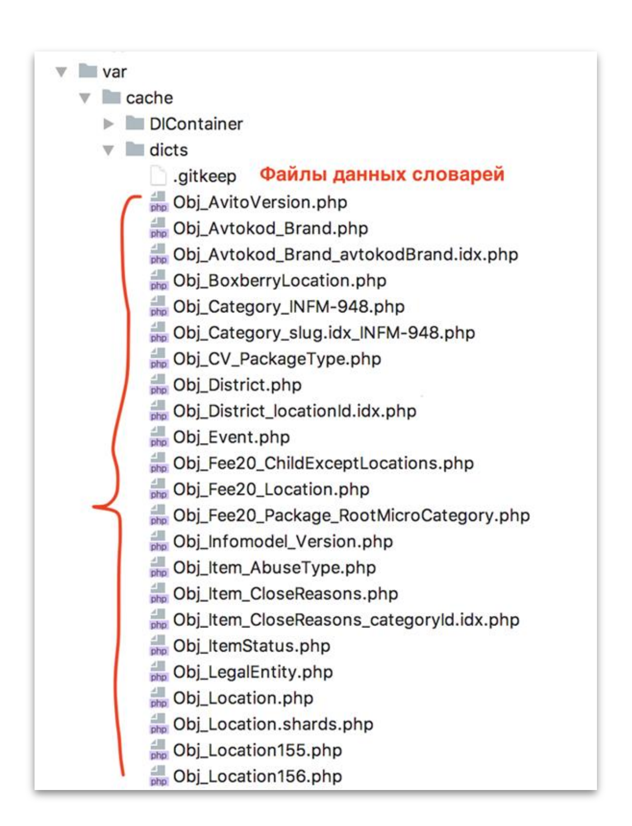
Описание иерархии

Add.php | MagicNumbersAdd.php | FirstLine.php | ParamValue.php | Obj_Location156.php

```php
<?php
// generated by Core\Utils\Obj\Dict::storeDict() (see Obj_Location::getDict())
return array (
  626740 =>
  array (
    'parentId' => 626470,
    'countryId' => 621540,
    'name' => 'Биробиджан',
    'nameGenitive' => 'Биробиджана',
    'namePrepositional' => 'Биробиджане',
    'slug' => 'birobidzhan',
    'type' => 3,
    'isCapital' => true,
    'regionCode' => 'R3',
    'id' => 626740,
  ),
  624360 =>
  array (
    'parentId' => 624300,
    'countryId' => 621540,
    'name' => 'Владимир',
    'nameGenitive' => 'Владимира',
    'namePrepositional' => 'Владимире',
    'slug' => 'vladimir',
    'type' => 3,
    'isCapital' => true,
    'regionCode' => 'R22',
    'id' => 624360,
  ),
  624840 =>
  array (
    'parentId' => 624770,
    'countryId' => 621540,
    'name' => 'Волгоград',
    'nameGenitive' => 'Волгограда',
    'namePrepositional' => 'Волгограде',
    'slug' => 'volgograd',
    'type' => 3,
```




Performance





Недостатки старой архитектуры

- Фундаментальные
- Наша реализация зависимостей
- Технические ограничения



Фундаментальный недостаток

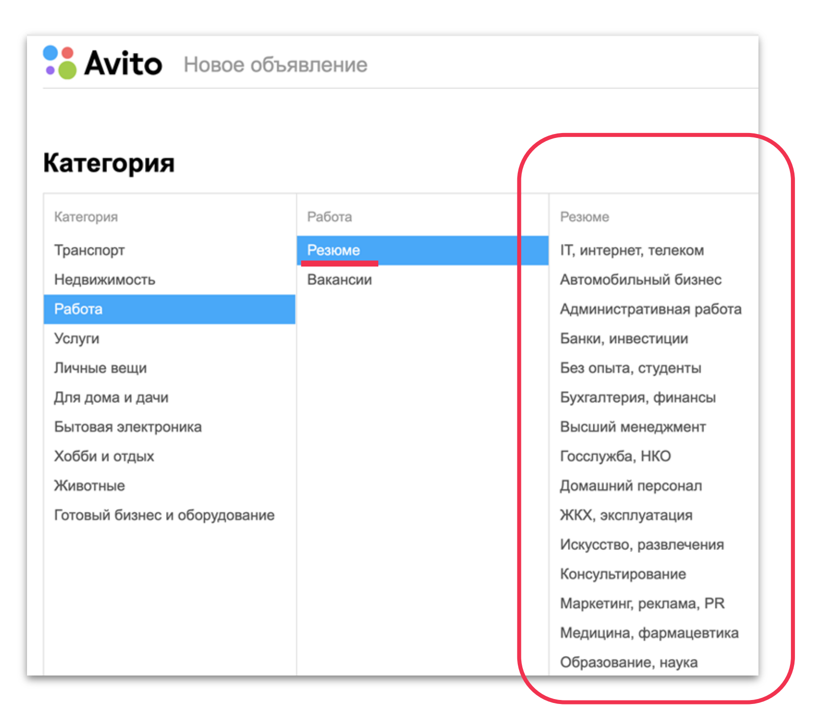
Изменение метаданных требует:

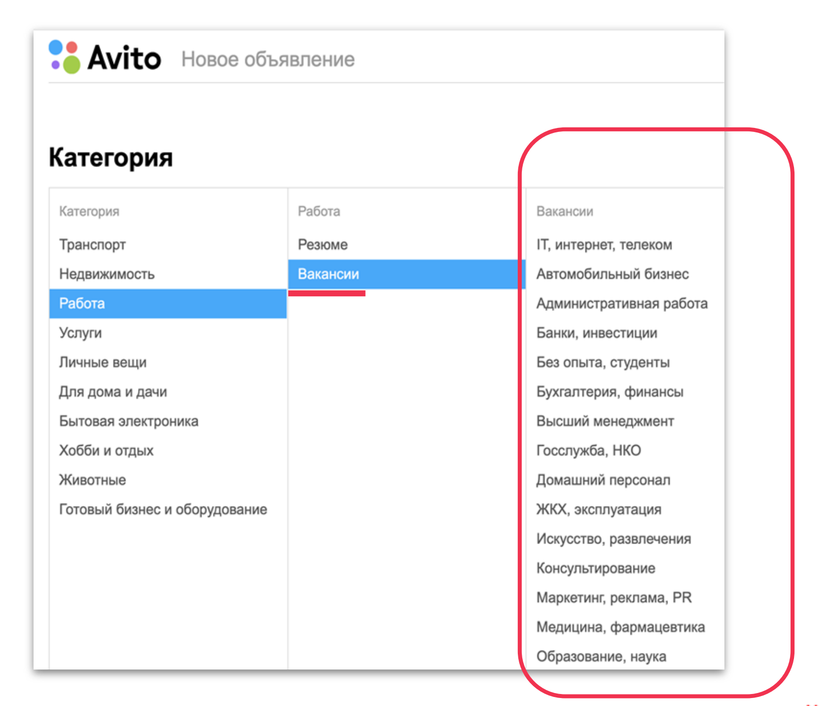
- Изменения схемы базы данных
- Пересборку проекта



Наша реализация зависимостей

Значения вакансий и резюме имеют разные идентификаторы:

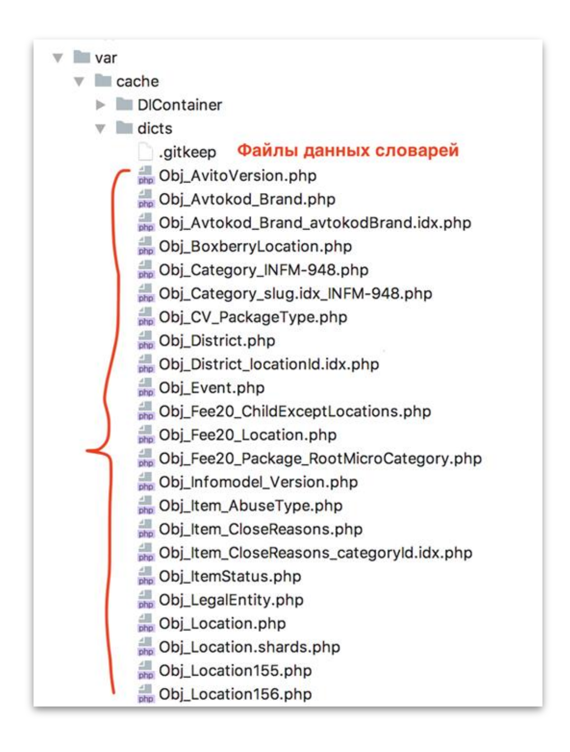






Технические ограничения

6k значений — это ~100 файлов словарей

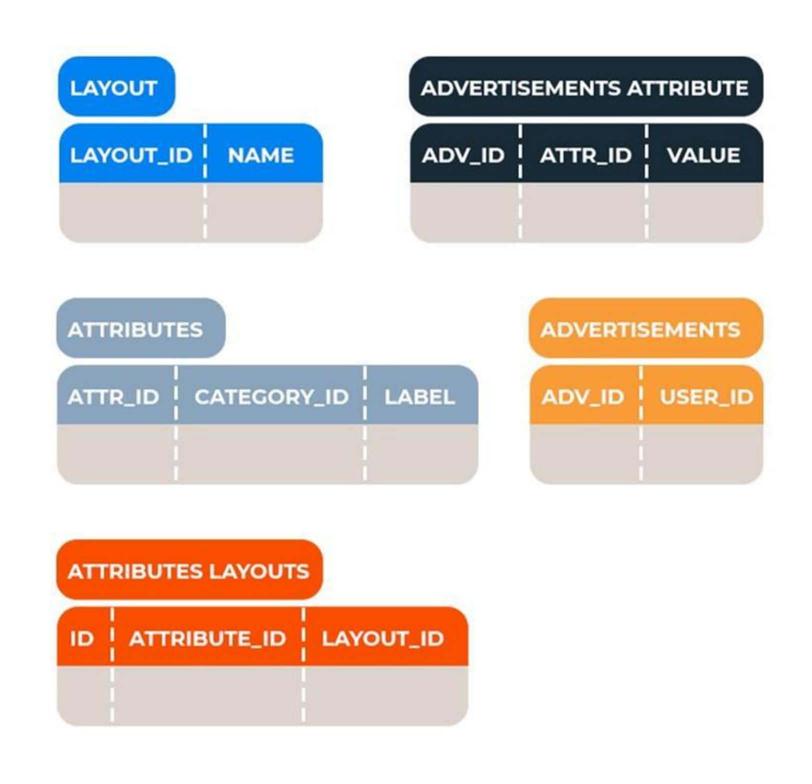


~10kk значений — это:

«Out of Memory Killer»



Entity Attribute Value модель





Functional requirements

- Уменьшить время внесения изменений
- Разное представление на разных платформах
- Поддержка a/b тестирования
- Поддержка версионирования

- **Управление метаданными**
- **Логическая консистентность**

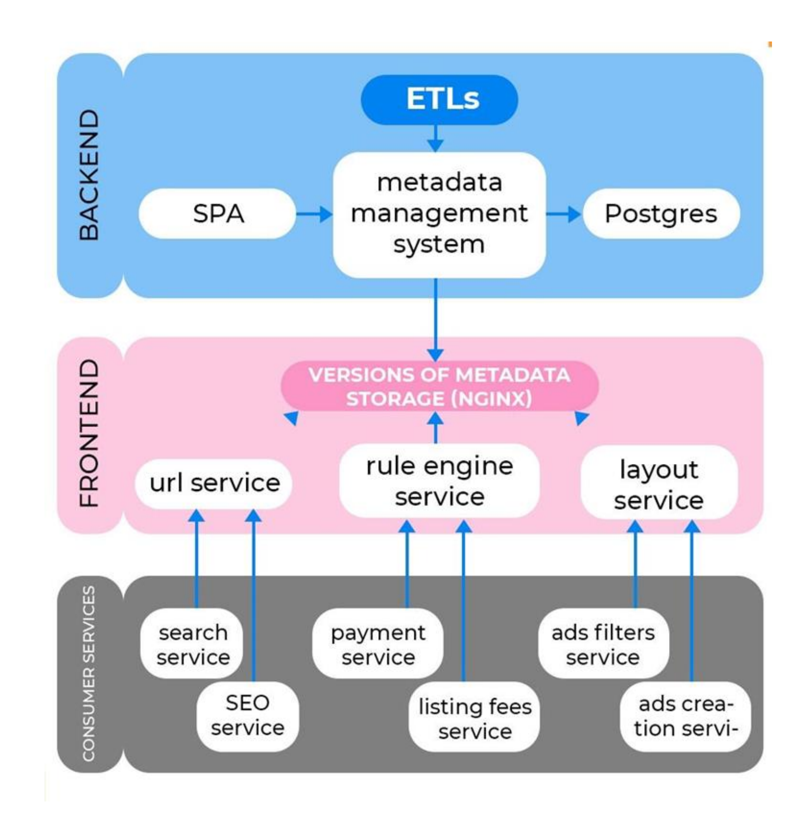


План

Архитектура
metadata
management system



Компонентная схема





Metadata management system

Categories / master / Infomode

Not Secure | infm.avito.ru/#/branches/master/categories/

Categories | Catalogs | Classification trees | Navigation | Forms | Branch main main

Infomodel / Categories

The branch was released by @mslanin а день назад (version INFM-3169). To make changes, create a new one.

Tests passed successfully | Release Description | Release logs

Categories

Search by category

AVITO	urls β	13 attributes	properties
Transport	urls β	0 attributes	properties
Cars	urls β	75 attributes	properties
Motorcycles and mototechnics	urls β	4 attribute	properties
Trucks and special equipment	urls β	fourteenattributes	properties
Water transport	urls β	3 attribute	properties
Parts and accessories	urls β	45 attributes	properties
The property	urls β	eleven attributes	properties
Apartments	urls β	106 attributes	properties
Rooms	urls β	68 attributes	properties
Houses, summer cottages, cottages	urls β	73 attribute	properties

infm.avito.ru/#/branches/master/trees/



Frontend слой



Data model

CATALOGS

Rules | Relations | Forms ← Search form

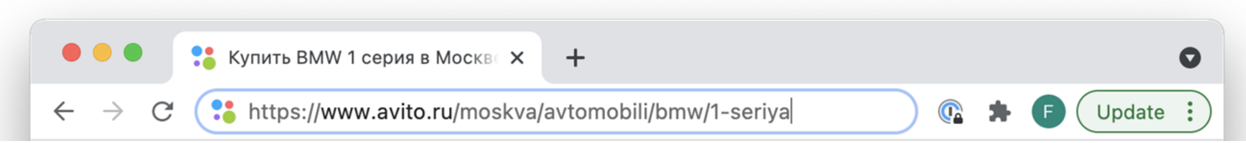
Rules | Relations | Forms ← Add form

Rules | Relations | Forms ← Card view

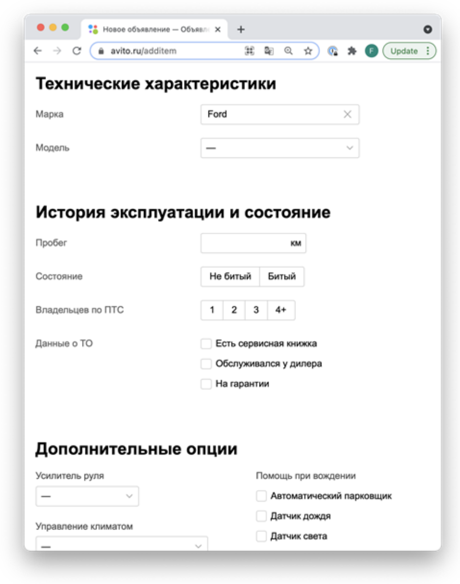


Примеры layouts

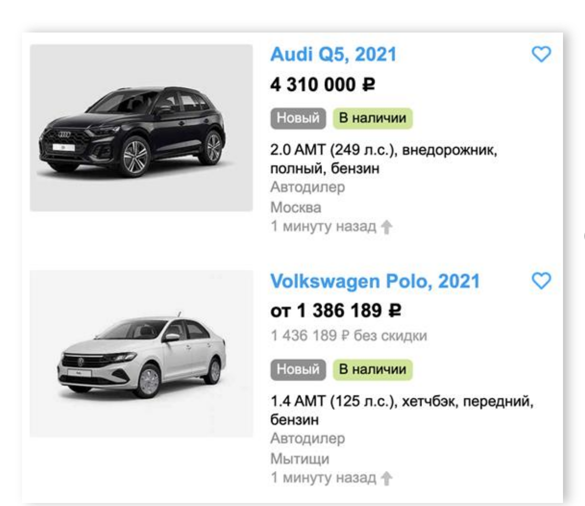
Canonical url



Формы



Сниппеты





Rules


```
- {
      field: 503,
    - states: [
        - {
              type: "visible",
            - constraints: [
                - [
                    - {
                          field: 201,
                        - condition: {
                            - values: [
                                  1060
                              ],
                              strict: false,
                          },
                      },
                    - {
                          field: 504,
                        - condition: {
                              empty: true
                          },
                      },
                  ],
                - [
                    - {
                          field: 201,
                        - condition: {
                            - values: [
                                  1060
                              ],
                              strict: false,
                          },
                      },
                    - {
                          field: 504,
                        - condition: {
                            - values: [
                                  5256
                              ],
                              strict: false,
                          },
                      },
                  ],
              ],
          }
      ],
  },
```




Relations

```
options: {
  - fields: [
    - {
        id: 110000,
        splitted: true,
      - values: [
        - {
          - value: {
                value: 329192,
                label: "AC",
            },
            fields: "relations/110000-329192.json.gz",
          },
        - {
          - value: {
                value: 329193,
                label: "Acura",
            },
            fields: "relations/110000-329193.json.gz",
          },
        - {
          - value: {
                value: 426153,
                label: "Adler",
            },
            fields: "relations/110000-426153.json.gz",
          },
        - {
          - value: {
                value: 329194,
                label: "Alfa Romeo",
            },
            fields: "relations/110000-329194.json.gz",
          },
        - {
          - value: {
                value: 329195,
                label: "Alpina",
            },
            fields: "relations/110000-329195.json.gz",
          },
        - {
          - value: {
                value: 473599,
                label: "Alpine",
            },
            fields: "relations/110000-473599.json.gz",
          },
        - {
          - value: {
                value: 427103,
                label: "AMC",
            },
            fields: "relations/110000-427103.json.gz",
```

```
{
    id: 110001,
  - values: [
    - {
      - value: {
            value: 329836,
            label: "CL",
        },
      - fields: [
        - {
            id: 110005,
          - values: [
            - {
              - value: {
                    value: 332068,
                    label: "I (1996–1999)",
                }
              },
            - {
              - value: {
                    value: 332069,
                    label: "II (2000–2003)",
                }
              },
            ],
          }
        ],
      },
    - {
      - value: {
            value: 329917,
            label: "CSX",
        },
      - fields: [
        - {
            id: 110005,
          - values: [
            - {
              - value: {
                    value: 332070,
                    label: "I (2005–2011)",
                }
              }
            ],
          }
        ],
      },
    - {
      - value: {
            value: 329990,
            label: "EL",
        },
      - fields: [
        - {
            id: 110005,
```




Forms


```
- fields: [
  - {
      id: 201,
      type: "enumerable",
      enumerable: true,
      dataType: "integer",
      slug: "offer_type",
      label: "Тип объявления",
    - validation: [
      - {
          rule: "required",
          message: "Выберите значение из списка",
          severity: "error",
        }
      ],
      multiple: false,
      range: false,
      fillable: true,
      searchable: true,
      immutable: true,
    - tags: [
        "semantic_41"
      ],
      defaultValue: null,
    - measureUnits: {
        one: null,
        two: null,
        five: null,
      },
      uid: 10774,
      name: "params[201]",
      placeholder: null,
      help: null,
    - widget: {
        type: "radioGroup",
      - config: {
          limit: 2,
          reduced: "true",
          withUrls: true,
          hideTitle: "true",
        },
      },
      sorting: 2,
      visible: true,
      stepId: 363,
      catalog: null,
      prevEntity: null,
    - parentEntity: {
        id: 363,
        type: "step",
      },
  },
  - {
```




Layout — гранулярность конфигурирования единых метаданных в разрезе платформы, категории, версии



Версионирование

Ветка master master

Поиск по веткам

All My AB Feature

AU-4081-parts-cars	feature	
JOB-4811	feature	
GEO-2556	ab	
master	master	
AU-4909	feature	
AU-4979-sort	feature	
AU-4594	feature	
AU-4894	feature	



Релиз версии метаданных

LAYOUT
LAYOUT_ID | NAME

ADVERTISEMENTS ATTRIBUTE
ADV_ID | ATTR_ID | VALUE

ATTRIBUTES
ATTR_ID | CATEGORY_ID | LABEL

ADVERTISEMENTS
ADV_ID | USER_ID

ATTRIBUTES LAYOUTS
ID | ATTRIBUTE_ID | LAYOUT_ID

Ветка JOB–4811 feature

Создать релиз и заблокировать ветку

nginx

версия job-4811

layout : desktop_filters

rules.json

relation.json

forms.json



a/b тестирование

LAYOUT
LAYOUT_ID | NAME

ADVERTISEMENTS ATTRIBUTE
ADV_ID | ATTR_ID | VALUE

ATTRIBUTES
ATTR_ID | CATEGORY_ID | LABEL

ADVERTISEMENTS
ADV_ID | USER_ID

ATTRIBUTES LAYOUTS
ID | ATTRIBUTE_ID | LAYOUT_ID

Ветка JOB–4811 feature

Создать релиз и заблокировать ветку

nginx

версия **ab-test-4811**

layout : desktop_filters

rules.json

relation.json

forms.json



Дистрибьюция релиза метаданных

INFOMODEL STORAGE
(NGINX)

version X
layout X
category X
category Y
category Z

version Y
layout X
category X
category Y
category Z

[layout, rule-engine,
url-builder] services

request [buld | validate | calc]
layout {version}.{layout}.
{category} for [data]



Routing

CATEGORY 1

layout advertisement_add	ver-1234.advertisement_add.cat1
layout advertisement_add	ver-1236.advertisement_add.cat1
layout advertisement_add	ver-1237.advertisement_add.cat1

CATEGORY 2

layout desktop_filter	ver-1234.desktop_filter.cat2
layout desktop_filter	ver-1236.desktop_filter.cat2
layout desktop_filter	ver-1237.desktop_filter.cat2



Демонстрация работы

```
{
    "route": {
        "categoryId": 24,
        "layout": "desktop_catalog_filters",
        "version": "master"
    },
    "form": [],
}
```

service-layout



Демонстрация работы

service-layout


```
{
    "result": {
        "form": [
            {
                "data": {
                    "id": 201,
                    "value": 1059,
                    "options": [
                        {
                            "id": 1059,
                            "label": "Купить"
                        },
                        {
                            "id": 1060,
                            "label": "Снять"
                        },
                        {
                            "id": 1058,
                            "label": "Покупатели"
                        },
                        {
                            "id": 1061,
                            "label": "Арендаторы"
                        }
                    ],
                    "label": "Тип объявления"
                },
                "needUpdate": true,
                "needDataSend": true
            },
            {
                "data": {
                    "id": 578,
                    "value": null,
                    "label": "Общая площадь"
                },
                "needUpdate": false,
                "needDataSend": false
            },
...

```




Демонстрация работы


```json
{
    "result": {
        "form": [
            {
                "data": {
                    "id": 201,
                    "value": 1059,
                    "options": [
                        {
                            "id": 1059,
                            "label": "Купить"
                        },
                        {
                            "id": 1060,
                            "label": "Снять"
                        },
                        {
                            "id": 1058,
                            "label": "Покупатели"
                        },
                        {
                            "id": 1061,
                            "label": "Арендаторы"
                        }
                    ],
                    "label": "Тип объявления"
                },
                "needUpdate": true,
                "needDataSend": true
            },
            {
                "data": {
                    "id": 578,
                    "value": null,
                    "label": "Общая площадь"
                },
                "needUpdate": false,
                "needDataSend": false
            },
...
```


Купить | Снять | ...

Количество комнат

- Студия
- Своб. планировка
- 1 комната
- 2 комнаты
- 3 комнаты
- 4 комнаты
- 5 комнат и более

Цена

Цена от | до, руб.

Вид объекта

Все | Вторичка | Новостройка

Общая площадь, м²

от | до

Этаж

от | до

- Не первый
- Не последний

Этажей в доме

от | до

Тип дома

- Кирпичный
- Монолитный



Демонстрация работы

Купить | Снять | ...

Количество комнат

- [ ] Студия
- [ ] Своб. планировка
- [ ] 1 комната
- [ ] 2 комнаты
- [ ] 3 комнаты
- [ ] 4 комнаты
- [ ] 5 комнат и более

Цена

Цена от | до, руб.

Вид объекта

Все | Вторичка | Новостройка

Срок сдачи

Срок сдачи

Общая площадь, м²

от | до

Этаж

от | до

- [ ] Не первый
- [ ] Не последний


```
{
    "route": {
        "categoryId": 24,
        "layout": "desktop_catalog_filters",
        "version": "master"
    },
    "form": [
        {
            "id": 499,
            "value": 5255
        },
...
```




Демонстрация работы

```
{
    "route": {
        "categoryId": 24,
        "layout": "desktop_catalog_filters",
        "version": "master"
    },
    "form": [
        {
            "id": 499,
            "value": 5255
        },
...
```

service-layout



Демонстрация работы

service-layout


```json
{
    "result": {
        "form": [
            {
                "data": {
                    "id": 110684,
                    "value": null,
                    "label": "Срок сдачи"
                },
                "needUpdate": false,
                "needDataSend": false
            },
            {
                "data": {
                    "id": 1383,
                    "value": null,
                    "label": "Название новостройки"
                },
                "needUpdate": false,
                "needDataSend": false
            },
            {
                "data": {
                    "id": 1634,
                    "value": null,
                    "label": "..."
                },
                "needUpdate": false,
                "needDataSend": false
            },
            {
                "data": {
                    "id": 100003,
                    "value": null,
                    "label": "Цена"
                },
                "needUpdate": false,
                "needDataSend": false
            }
        ]
    }
}
```




Демонстрация работы


```json
{
    "result": {
        "form": [
            {
                "data": {
                    "id": 110684,
                    "value": null,
                    "label": "Срок сдачи"
                },
                "needUpdate": false,
                "needDataSend": false
            },
            {
                "data": {
                    "id": 1383,
                    "value": null,
                    "label": "Название новостройки"
                },
                "needUpdate": false,
                "needDataSend": false
            },
            {
                "data": {
                    "id": 1634,
                    "value": null,
                    "label": "..."
                },
                "needUpdate": false,
                "needDataSend": false
            },
            {
                "data": {
                    "id": 100003,
                    "value": null,
                    "label": "Цена"
                },
                "needUpdate": false,
                "needDataSend": false
            }
        ]
    }
}
```


Купить | Снять | ...

Количество комнат

- Студия
- Своб. планировка
- 1 комната
- 2 комнаты
- 3 комнаты
- 4 комнаты
- 5 комнат и более

Цена

Цена от | до, руб.

Вид объекта

Все | Вторичка | Новостройка

Срок сдачи

Срок сдачи

—

Дом сдан

2021

до 2 квартала 2021

до 3 квартала 2021

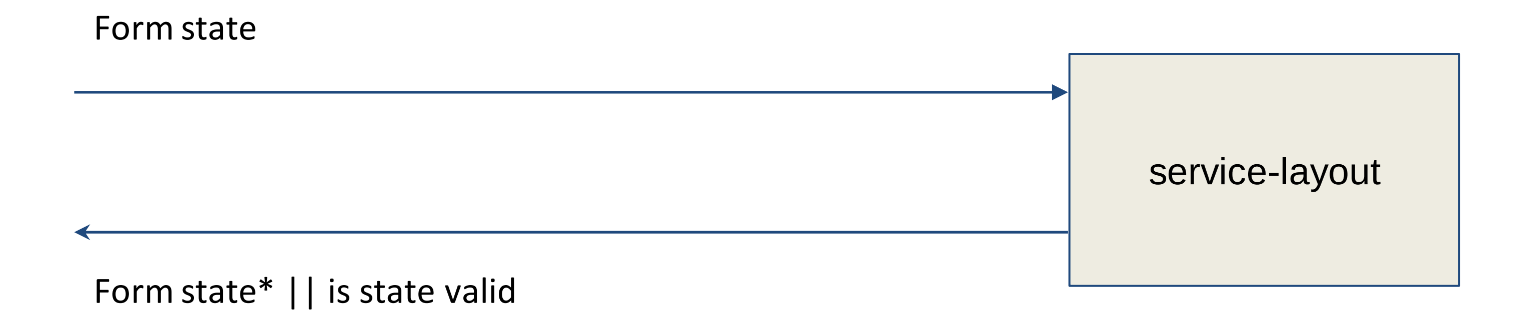
до 4 квартала 2021

2022

до 1 квартала 2022



Получение следующего состояния





План

Deep dive
сервиса **layout**



Non functional requirements

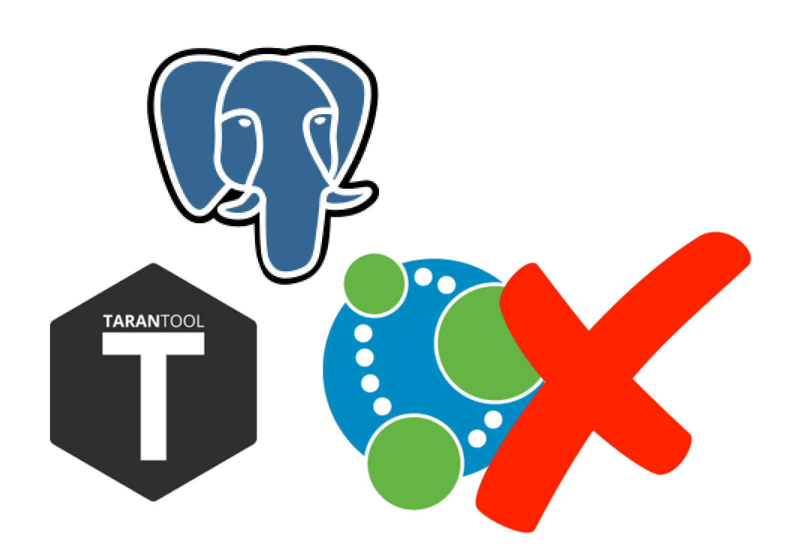
- Response time: 10ms на 99%%
- Должен держать в памяти

  <~10 000 layouts

  (layout может занимать до 500Mb)



Выбор хранилища







Request flow diagram

Consumer service

[layout, rule-engine, url-builder] services

service

interpreter

in memory cache

"compiler"

loader

Infomodel Storage

request [build|validate]
layout {**version**}.{**layout**}.{**category**}
for [data]

see if layout {**version**}.{**layout**}.
{**category**} is already compiled and in
memory

No required layout compiled

Interpret data

response

miro



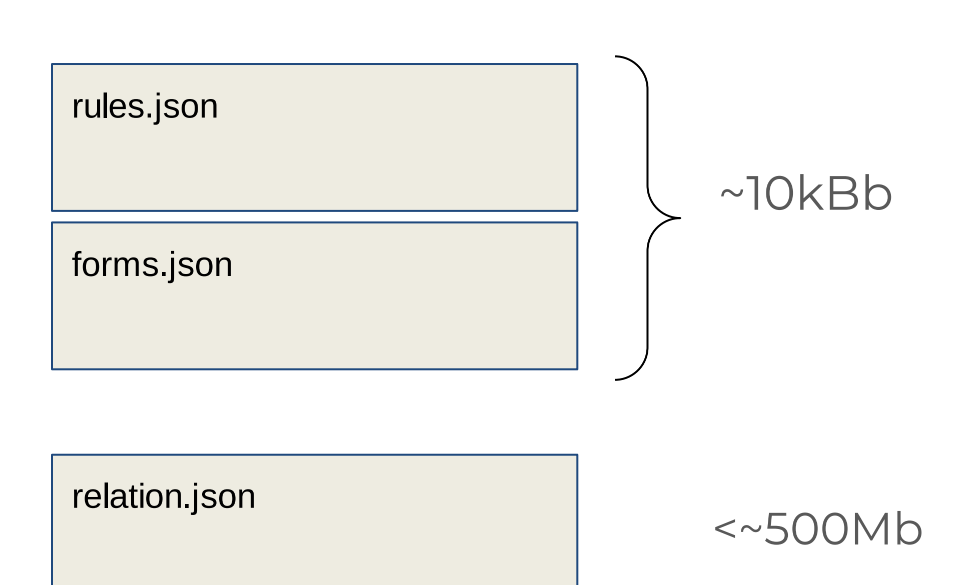
Проблема хранения данных

В худшем случае количество памяти для 1500 layouts - 700Gb!

На один инстанс сервиса



Внутренняя структура хранилища





Пример дерева зависимостей

localhost

```
$ match p=(n:fieldTree)-[:rels]->()-[:rels]->()-[:rels]->()-[:rels]->()-[:rels]->()-[:rels]->()-[:rels]->()-[:rels]->()-[:rels]->()-[:rels]->() where n.fieldId=110000 return p
```

*(300) fieldTree(300)

*(299) rels(299)

Graph
Table
Text
Code

Displaying 300 nodes, 299 relationships.



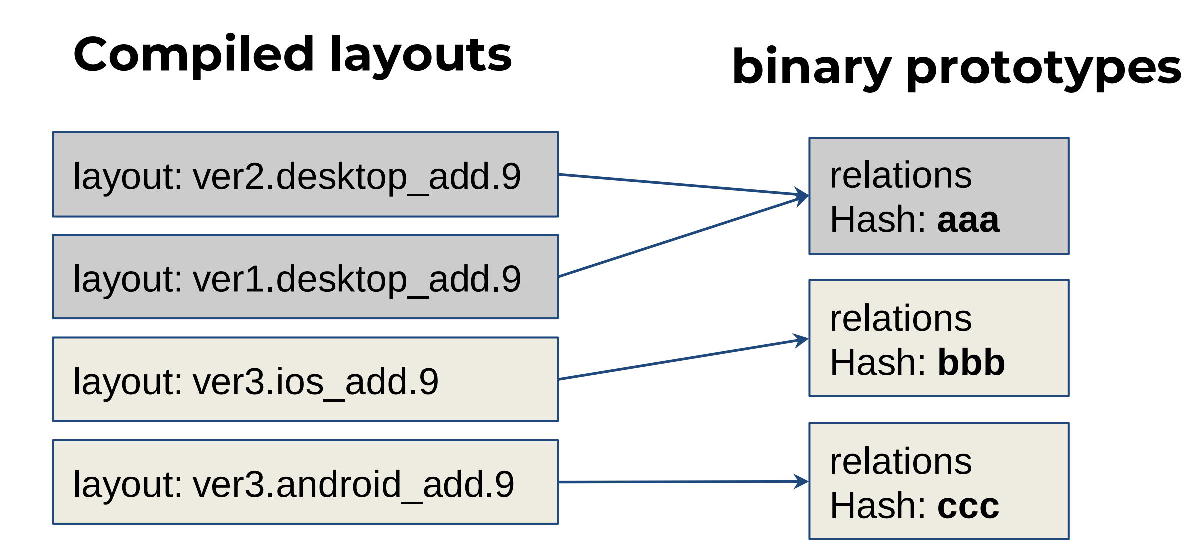
Переиспользование прототипов

50Mb — все еще

много



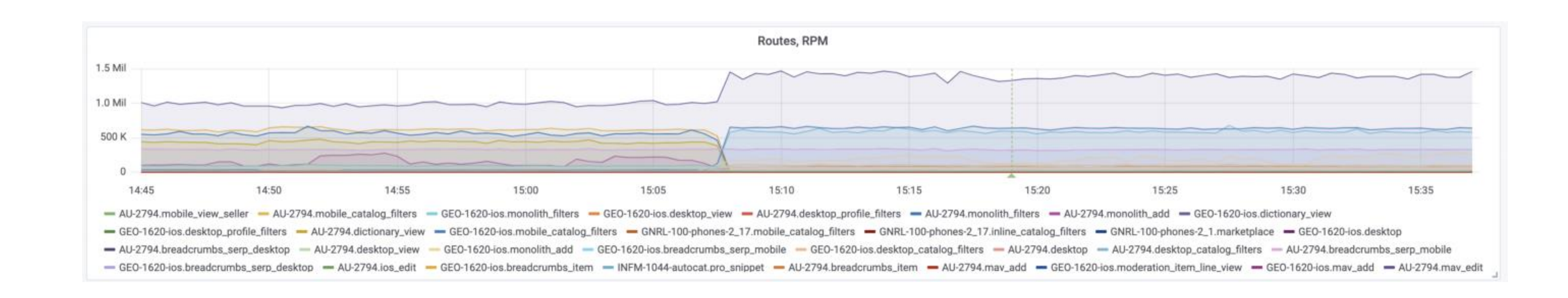
Переиспользование прототипов





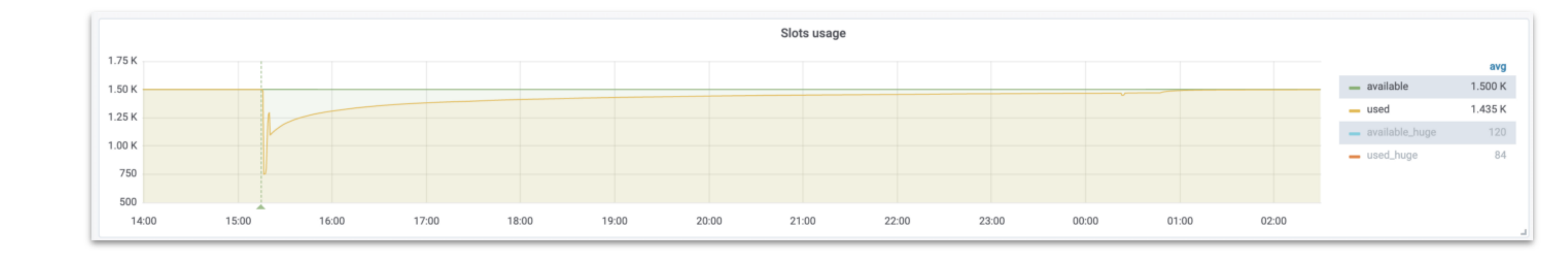
Политики вытеснения

LRU - x; LFU - x; ARC -v





Политика вытеснения layouts





Cache warmup

NGINX

download actual
layout specs

read the list of layouts
that are in the fly

write the current list of
layouts in in-memory cache

Instances of
Service that
spinning up

layout
service
instance

layout
service
instance

Instances of
Service
in production

miro



План

Плюсы/
минусы



Плюсы

- Быстрый релиз изменений
- Горизонтальное масштабирование системы
- Расширяемая архитектура



Минусы

- Значительное усложнение архитектуры
- Высокое потребление сети
- Высокое потребление cpu



Спасибо за внимание

статья на medium



